Issues and challenges for MME systems

Huug van den Dool

Climate Prediction Center (CPC)

Acknowledge: Emily Becker, Qin Zhang, Suranjana Saha, Malaquias Pena and many others.

La Jolla, July 30, 2013

In Early 2011

- There was a sudden decisiveness about organizing a National MME for seasonal prediction in the US
- It had been a longstanding wish of some, especially funding agents, for this to happen.
- In a sense, we were ready, since IMME was already being prepared.
- There was a willingness to go the extra mile on the part of other modeling centers, especially NASA, GFDL, NCAR and IRI to get this done quickly.
- These were all global coupled atmosphere-ocean models.
- NCEP organized the “rules of engagement” such as time table, common grid, hindcasts, etc.
- The first test run in real time was made in August 2011.

Requirements for NMME (Huug van den Dool, 4/7/11)

- Real-time model should be the frozen hindcast model (of course, the initial states may change due to ingest of new data types)
- It would be good to follow the CFSv2 lay-out for the start times of the hindcasts.
- Forecast leads out to at least 9 months.
- A minimum of 30 years of hindcasts, especially a common period of 1981-2010.
- The number of ensemble members is up to the originator, but it is assumed they understand that one ensemble member will keep skill low, and for very large N, the "threshold" returns are diminishing, so they have to make a wise choice.
- All individual members must be submitted, not just the ensemble mean.
- Total fields (not anomalies) must be submitted with no systematic error correction at the originator's end.
- Resolution and physics/numerics of the model are left entirely up to originators.
- Required output would minimally be monthly means of global SST, T2m, prate (in the first instance). Recently, T_{min} T_{max} , runoff, soil moisture and 200 hPa geopotential have been added.
- All data must be submitted in a common 1 x 1 degree grid.
- What about real time operations? All forecasts must be in by the 8^{th} of the month , so that they can be used as a tool in CPC's official seasonal predictions.

Hindcast Situation YEAR 1

Model resident Resolutions

	Start months available NOW	Period	Members	Arrangement of Members	Lead (months)	Atmosphere	Ocean	Reference	
NCEP-CFSv1	12	1981-2009	15	1st 0Z +/-2days, 11th 0Z+/-2d, 21st 0Z+/-2d	0-9	T62L64	MOM3L40 0.30 deq Eq	Saha et al 2006	NCEP-CFSv1
NCEP-CFSv2	12	1982-2010	24(28)	4 members (0,6,12,18Z) every 5th day	0-9	T126L64	MOM4 L40 0.25 deg Eq	Saha et al 2012	NCEP-CFSv2
GFDL-CM2.1	12	1982-2010	10	All 1st of the month 0Z	0-11	2x2.5deg L24	MOM4 L50 0.25 deg Eq	Delworth et al 2006	GFDL-CM2.1
IRI-Echam4-f	12	1982-2010	12	All 1st of the month**	0-7	T42L19	MOM3 L25 0.5 deg Eq	DeWitt MWR2005	IRI-Echam4-f
IRI-Echam4-a	12	1982-2010	12	All 1st of the month**	0-7	T42L19	MOM3 L25 0.5 deg Eq	"	IRI-Echam4-a
NCAR-CCSM3.0	12	1982-2010	6	All 1st of the month**	0-11	T85L26	POP L40 0.3 deg Eq	Kirtman and Min 2009	NCAR-CCSM3.0
NASA	12	1981-2010	6	1 member every 5th day as CFSv2	0-9	1x1.25deg L72	MOM4 L40 0.25 deg Eq	Rienecker et al 2008	NASA

Hindcast Situation YEAR 2

	Start months available NOW	Period	Members	Arrangement of Members	Lead (months)	Model resident Resolutions: Atmosphere	Ocean	Reference	
NCEP-CFSv1	12	1981-2009	15	1st 0Z +/-2days, 11th 0Z+/-2d, 21st 0Z+/-2d	0-9	T62L64	MOM3L40 0.30 deq Eq	Saha et al 2006	NCEP-CFSv1
NCEP-CFSv2	12	1982-2010	24(28)	4 members (0,6,12,18Z) every 5th day	0-9	T126L64	MOM4 L40 0.25 deg Eq	Saha et al 2010	NCEP-CFSv2
GFDL-CM2.1	12	1982-2010	10	All 1st of the month 0Z	0-11	2x2.5deg L24	MOM4 L50 0.30 deg Eq	Delworth et al 2006	GFDL-CM2.1
CMC1-CanCM3	12	1981-2010	10	All 1st of the month 0Z	0-11	CanAM3 T63L31	CanOM4 L40 0.94 deg Eq	Merryfield et al 2012	CMC1
CMC2-CanCM4	12	1981-2010	10	All 1st of the month 0Z	0-11	CanAM4 T63L35	CanOM4 L40 0.94 deg Eq	Merryfield et al 2012	CMC2
NCAR-CCSM3.0	12	1982-2010	6	All 1st of the month**	0-11	T85L26	POP L40 0.3 deg Eq	Kirtman and Min 2009	NCAR-CCSM3.0
NASA	12	1981-2010	6	1 member every 5th day as CFSv2	0-9	1x1.25deg L72	MOM4 L40 1/4 deg at Eq	Rienecker et al 2008	NASA

Issues

- What happened in real time? Aug 2011 forward
- Live Performance in Yr 1 and Yr 2
- Bias error *
- Role of soil moisture (indirectly) *
- Soil Moisture initialization *
- Prediction and Predictability
- (The upward trend in SST and T2m)*

* Paper in CDPW, College Park October 2013

Real Time Operations

- Only on one occasion was one model completely missing in real time
- On several occasions, one or more centers had “last minute” contributions
- Some centers had a smaller than intended number of ensemble members when they ran out of time.
- At least once a model was ‘wrong’. Undetected at furst, below the QC radar. The corrected version was later put into the archive.

Real Time Operations (contd)

- QC is badly needed. Looking at maps will do a lot.
- We assumed that quality control is typically about outliers.
- Missing members, all undefined members do happen
- Names can change suddenly
- However, quite often, members i and j of some models are erroneously identical, which is very unexpected !!

Real Time Operations (contd)

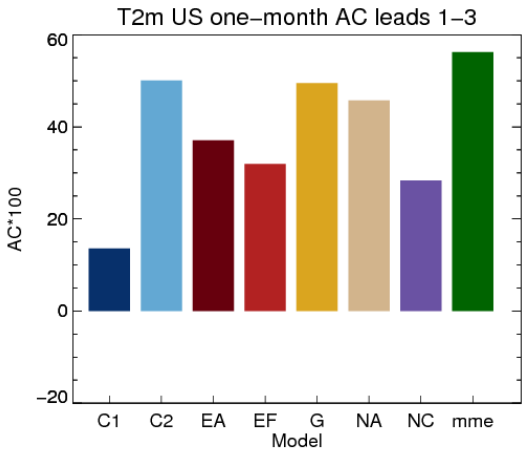
- Center X had to move its model to a new platform. No problem was anticipated, so they did not feel the need to inform anybody.
- However, very large positive T2m anomalies crept up in summer forecasts (at all leads), which was rather suspicious. Center X was unable to reproduce the real time forecast of the older model.
- They had to redo all hindcasts to match the 'new' model.
- On the positive side: For Center X, we now have two complete sets of hindcasts (all 12 start months) 1981-2011 which can be used for research at least.

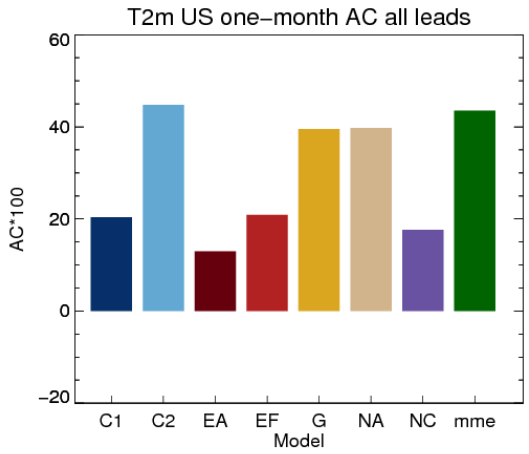
Real Time Operations (contd)

- Center Y decided that its soil initialization was not good enough and changed that aspect of initialization in May 2012.
- They continued with hindcasts on the fly.
- We had to wait a whole year to get the updated hindcasts complete.
- Models could keep changing in this fashion, which is frustrating, but reality, and manageable only if you want it badly enough.

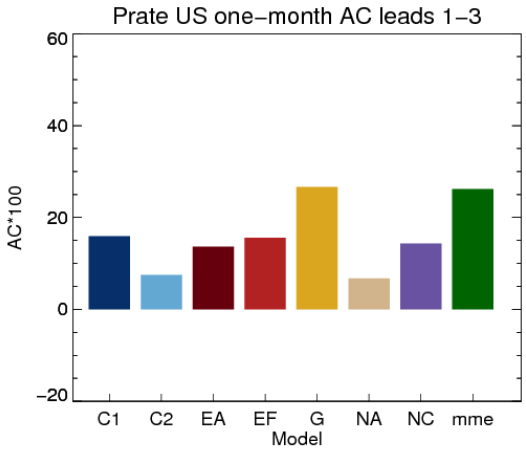
Year 1 One-month averages by lead*

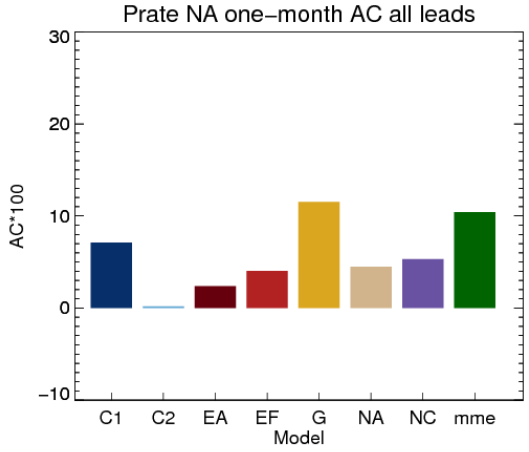
T2m





Prate

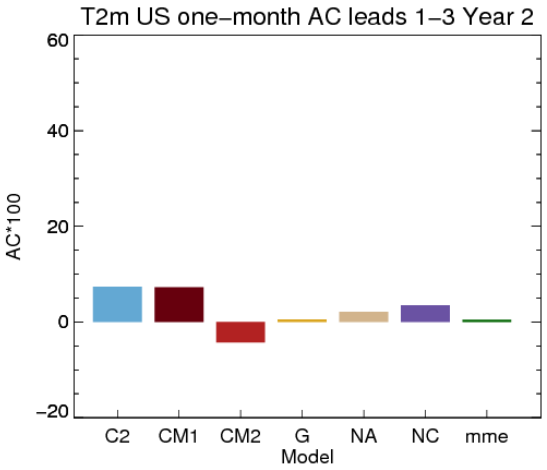


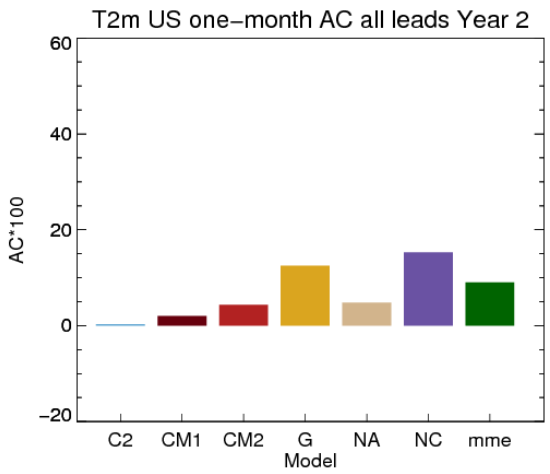


* Calculated for August 2011 – July 2012 initial conditions

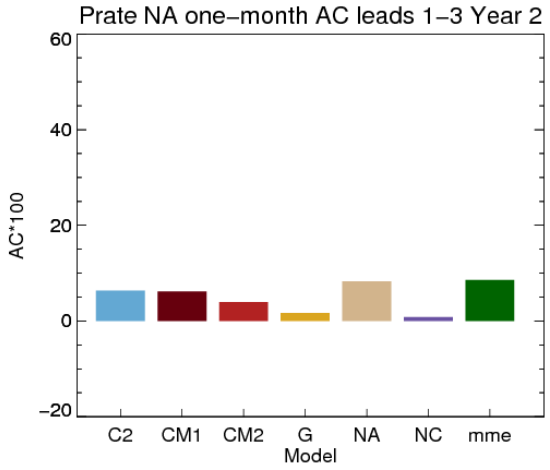
Year 2 One-month averages by lead*

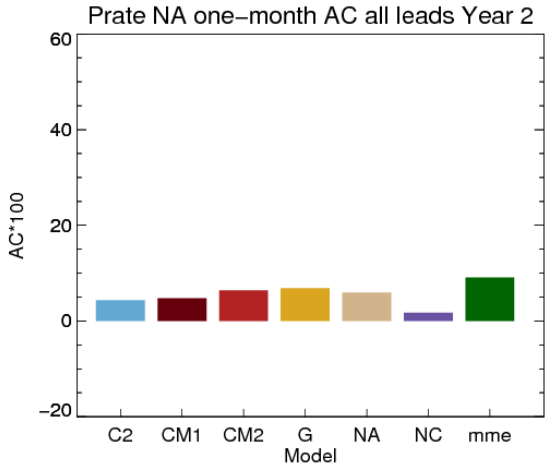
T2m





Prate

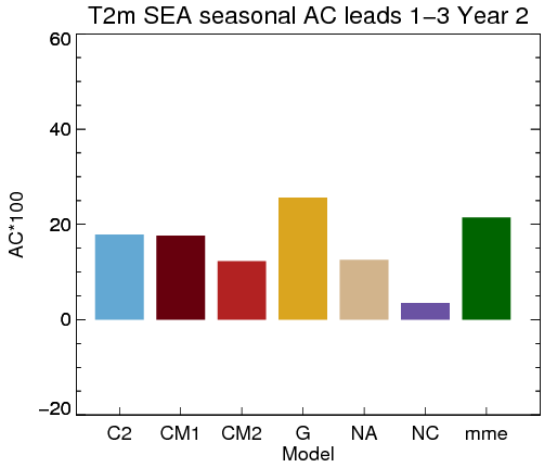


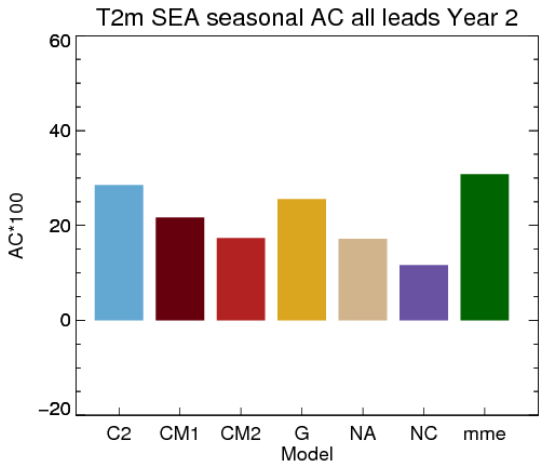


* Calculated for August 2012 – May 2013 initial conditions

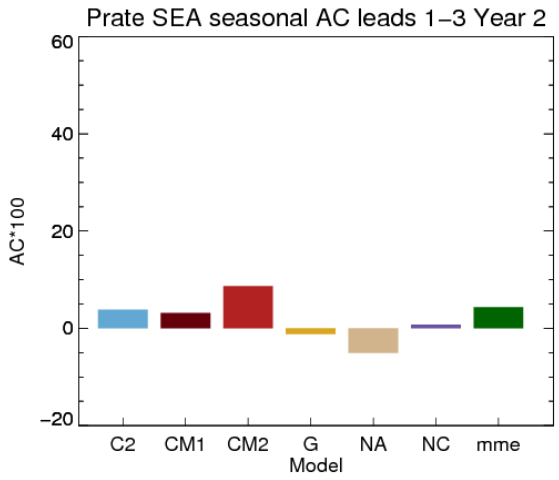
Year 2 Three-month averages by lead*

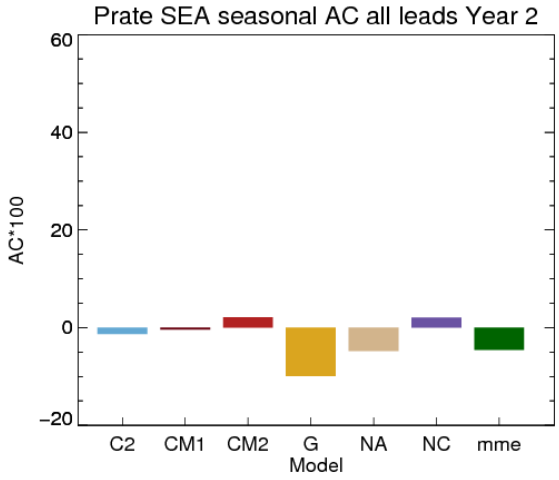
T2m





Prate



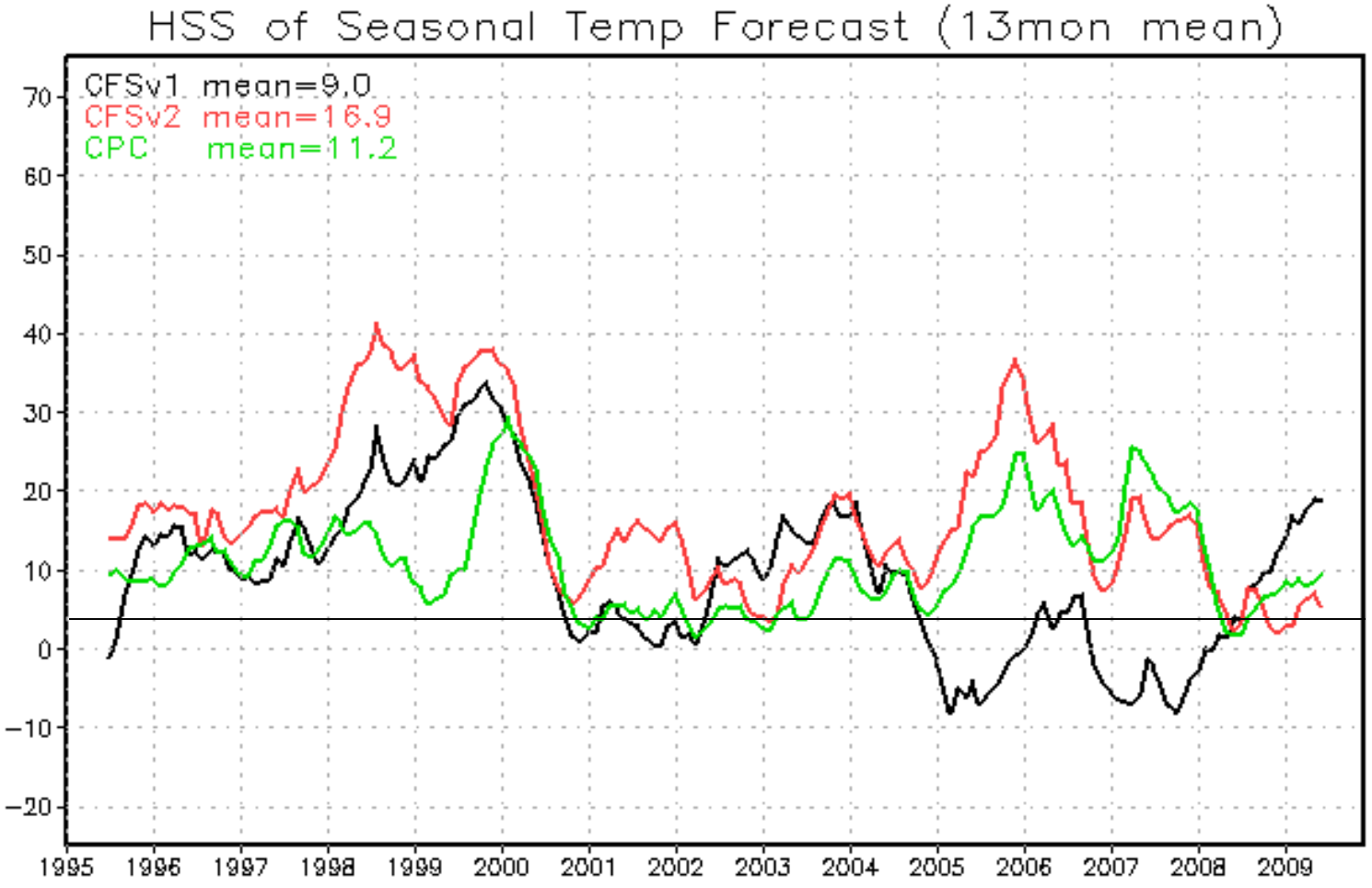


South East Asia
5N-50N
70E-135E

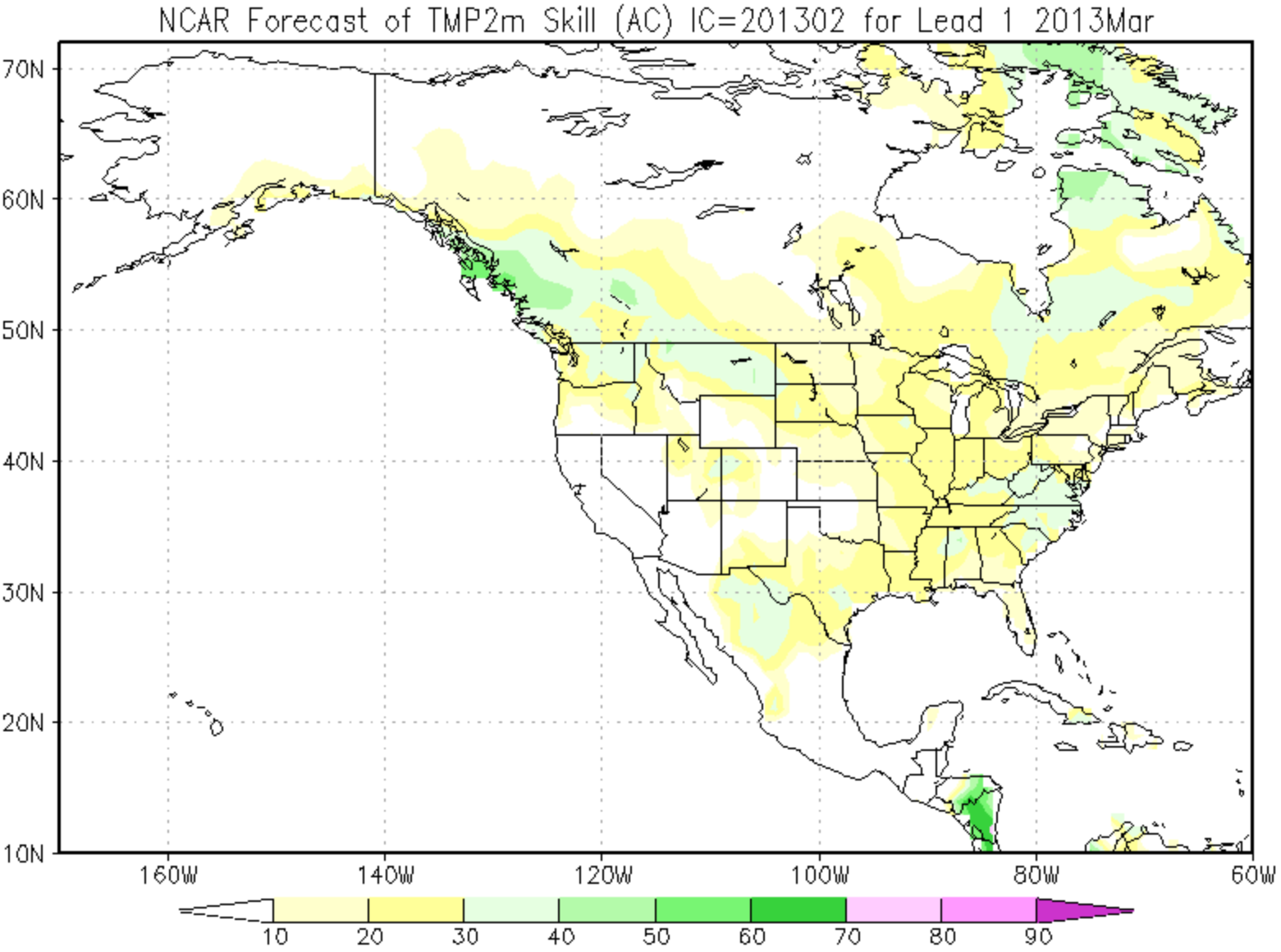
* Calculated for August 2012 – May 2013 initial conditions

Heidke Skill Score for 2-meter Temp

More skill for CFSv2



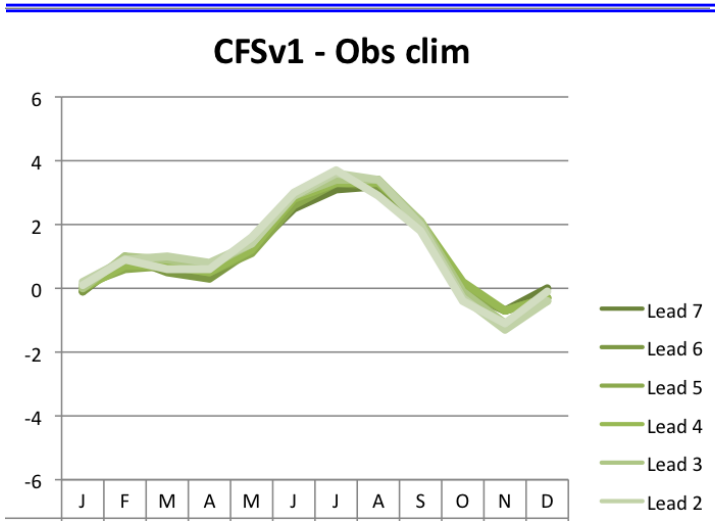
Peitao Peng, CPC

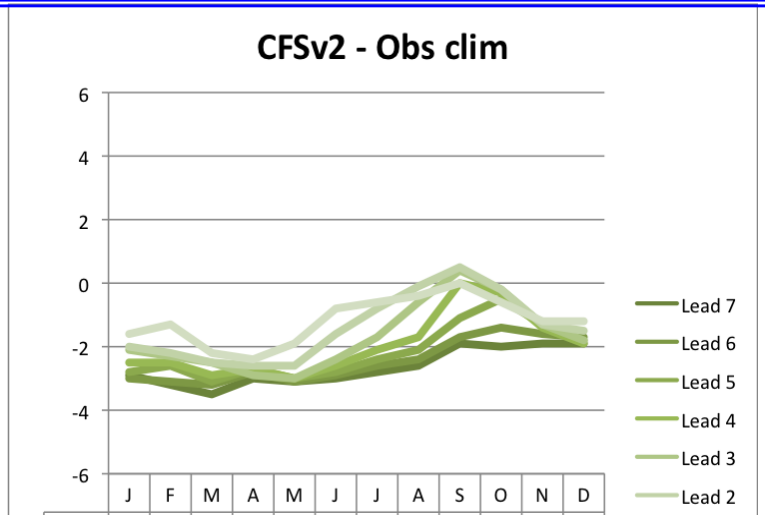


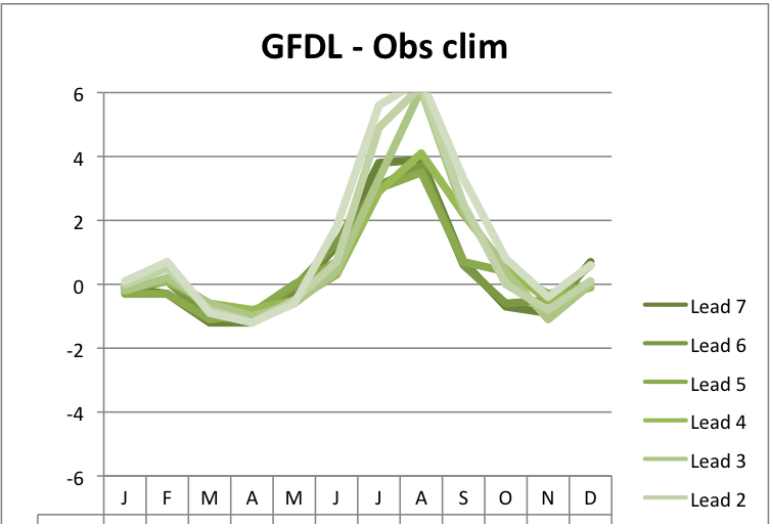
The a-priori skill estimate for March 2013 is based on the corresponding hindcasts 1982-present.

About the bias

Verification of Climate Mean

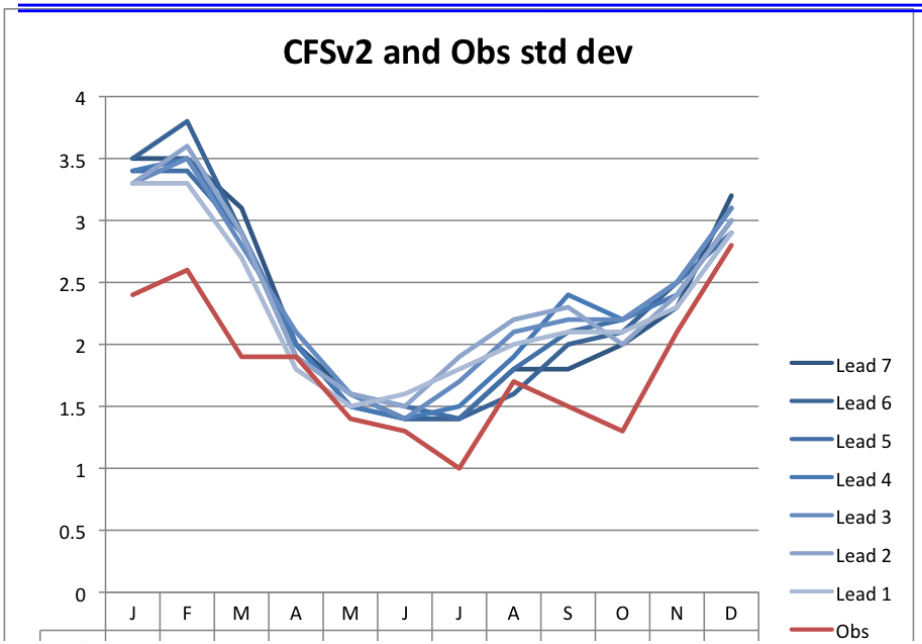


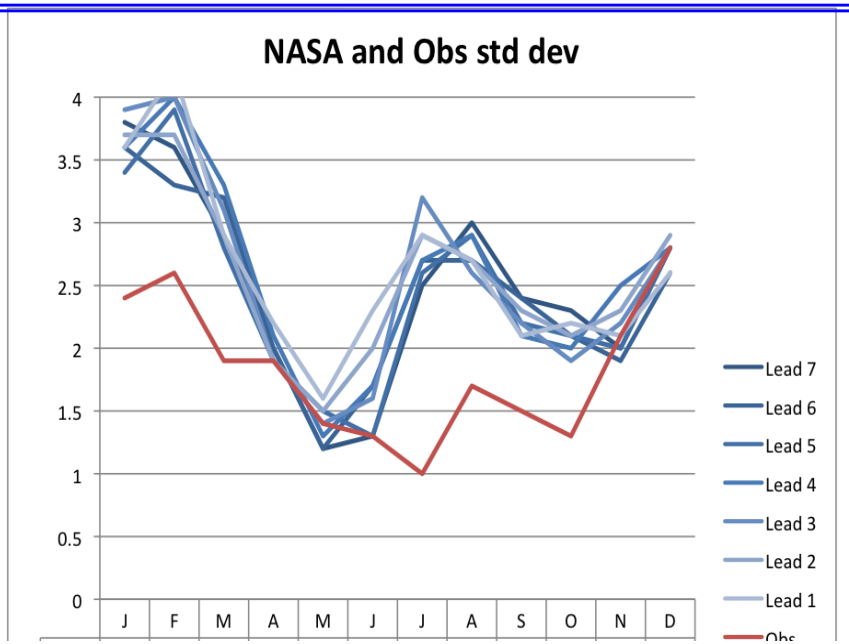




Points to note:
Target is more important than lead
CFSv2 is the only model with cold bias (center US)
Many models have (up to horrible) warm bias in summer climatology

Verification of Climate Variance





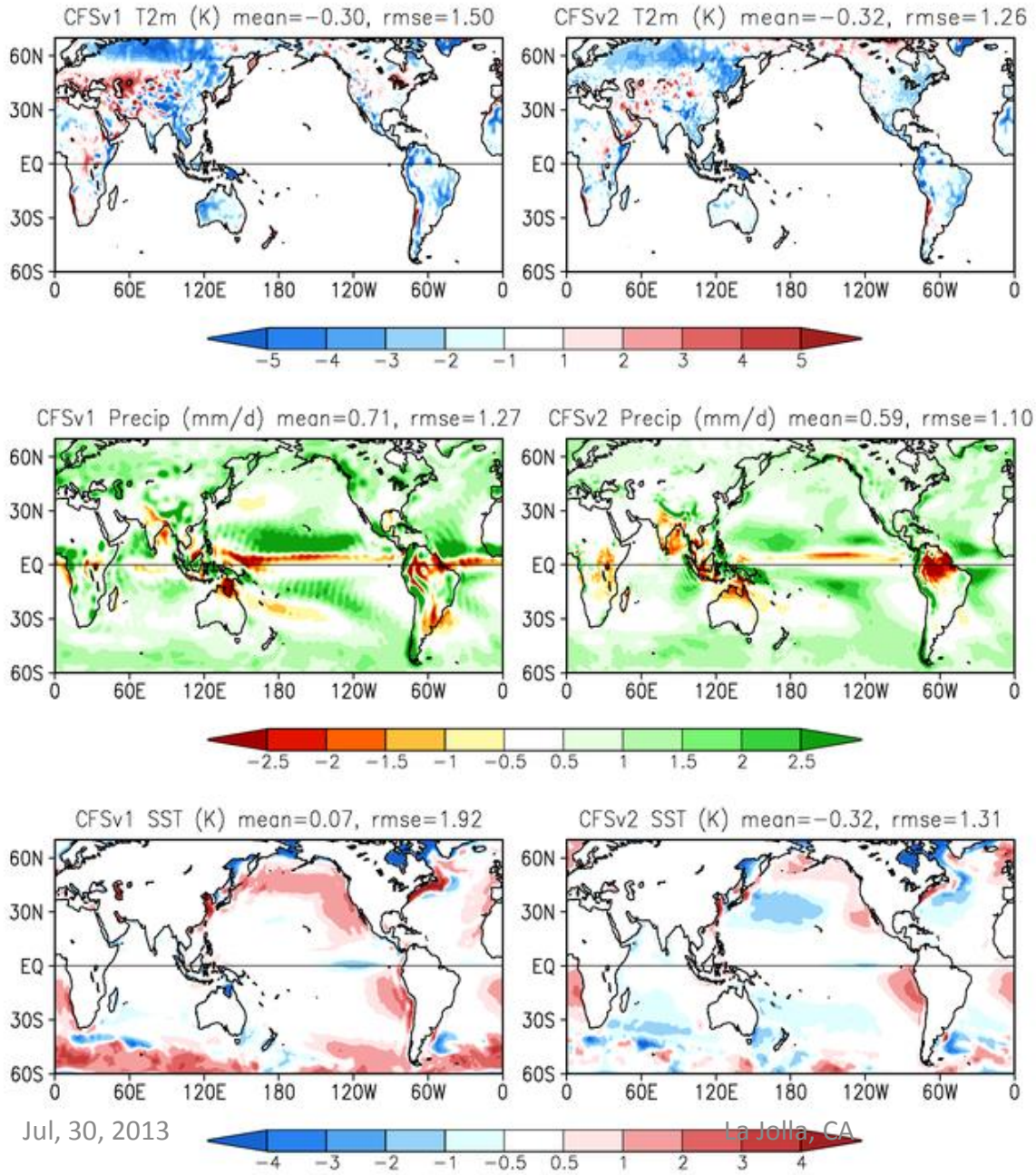
Points to note:

CFS has a reasonable seasonality in interannual standard deviation of monthly T2m (center US)

Several model have a spurious maximum in summer (NASA, GFDL). Reasons can be given off line

If anything, all models have too much spread thruout the year.

Annual Mean Systematic Error (1982–2009) for Lead 3



GHCN-CAMS (validation for Tmp2m)
CMAP (validation for Prate)
OIv2 (validation for SST)

Monthly lagged precip-temperature relationship in NMME.

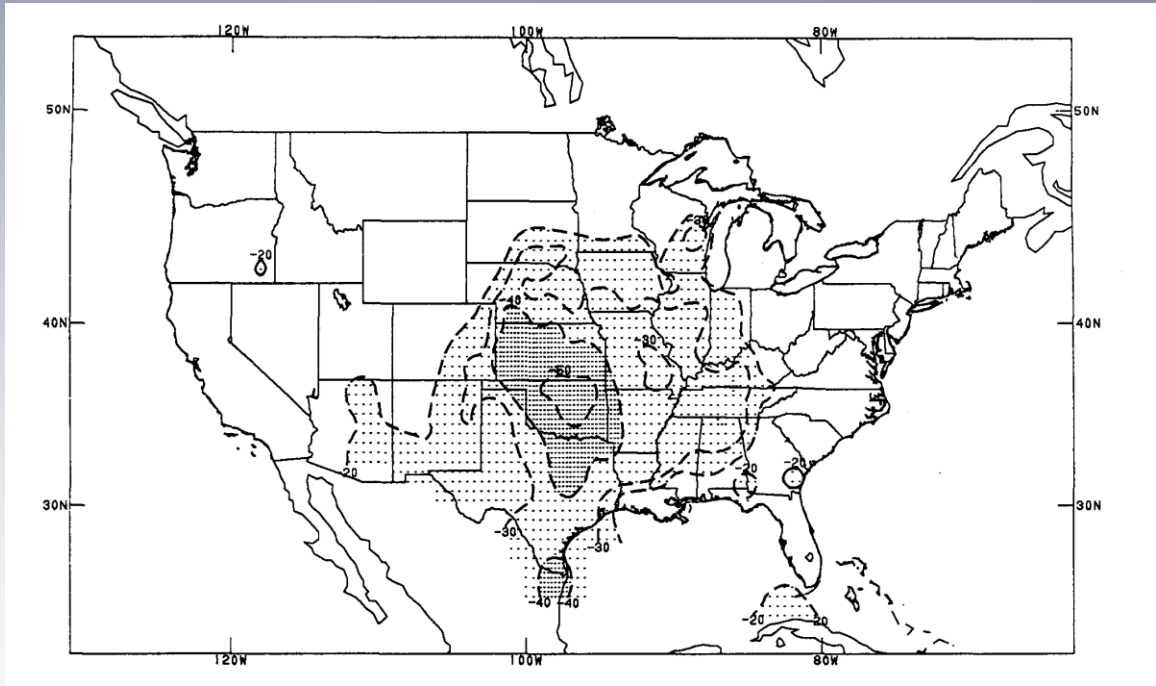
To gauge the activity of soil hydrology

Emily Becker & Huug van den Dool

NMME telecon, May 2

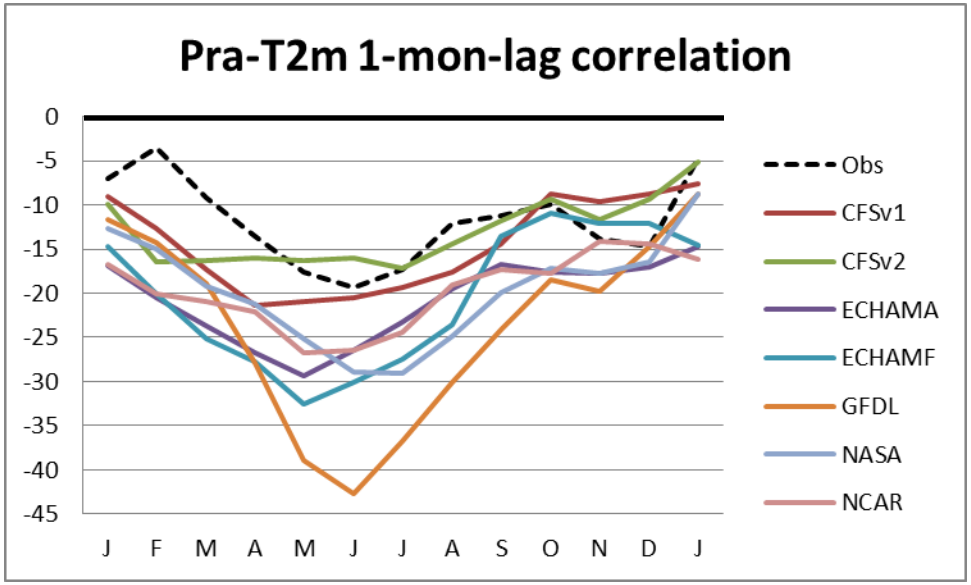
- 1-mon lagged correlation between precip and temp turns out to be negative
- Dry July → warm Aug; Wet July → cool Aug

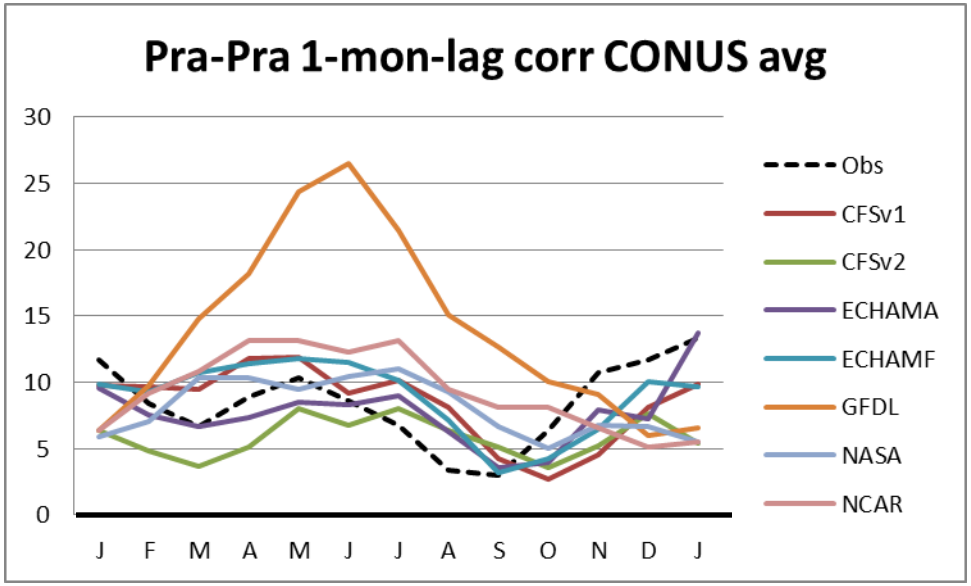
CD data 1931-1987

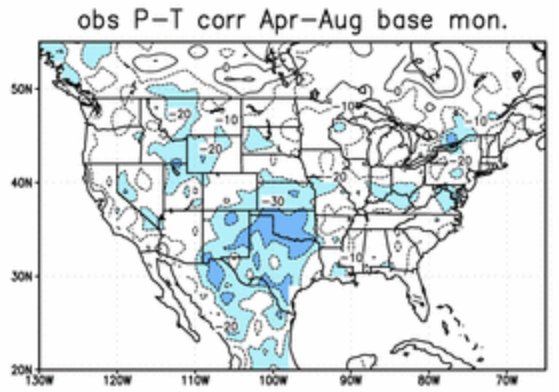


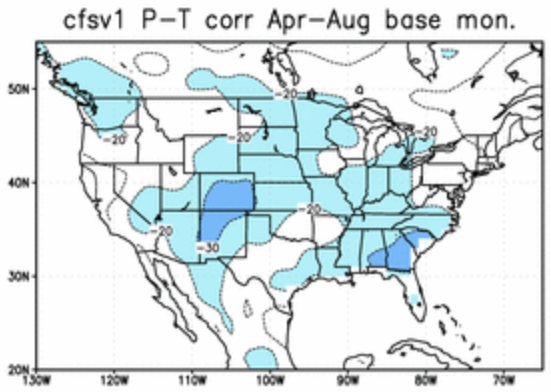
Monthly Precipitation–Temperature Relations and Temperature Prediction over the United States

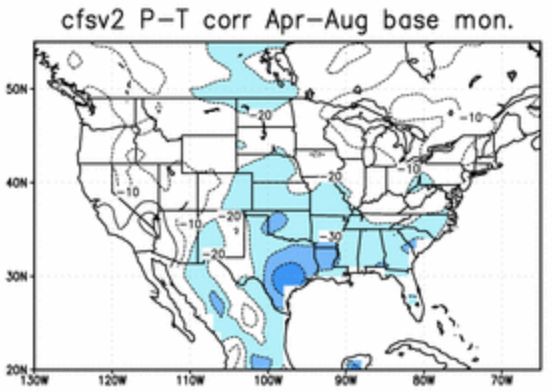
JIN HUANG AND HUUG M. VAN DEN DOOL

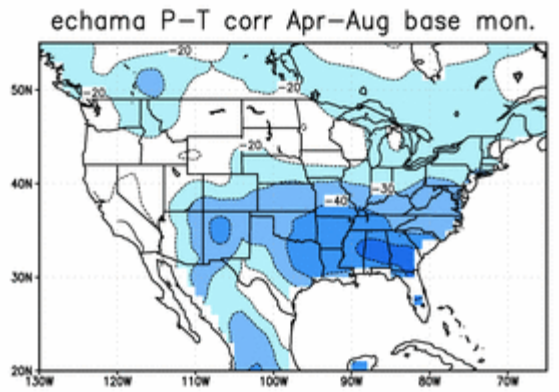

Pra-T2m 1-mon-lag correlation
0
-5
-10
-15
-20
-25
-30
-35
-40
-45
J F M A M J J A S O N D J
Obs
CFSv1
CFSv2
ECHAMA
ECHAMF
GFDL
NASA
NCAR

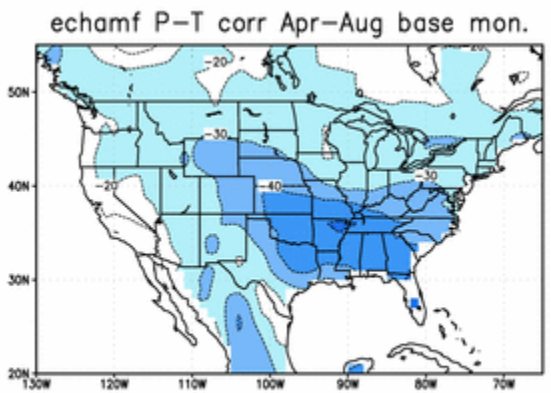

Pra-Pra 1-mon-lag corr CONUS avg
30
25
20
15
10
5
0
J F M A M J J A S O N D J
Obs
CFSv1
CFSv2
ECHAMA
ECHAMF
GFDL
NASA
NCAR

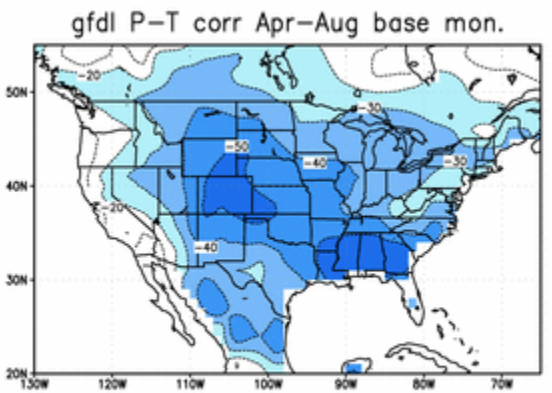


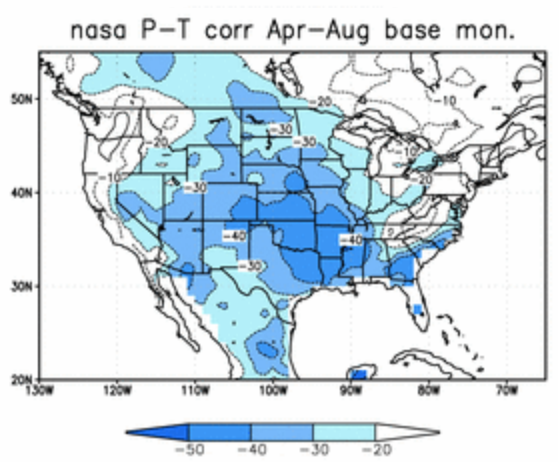


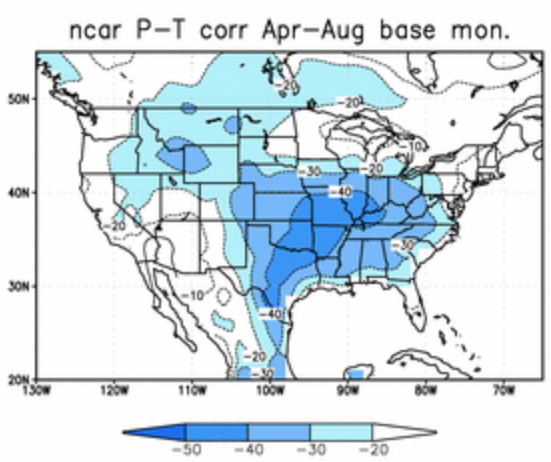










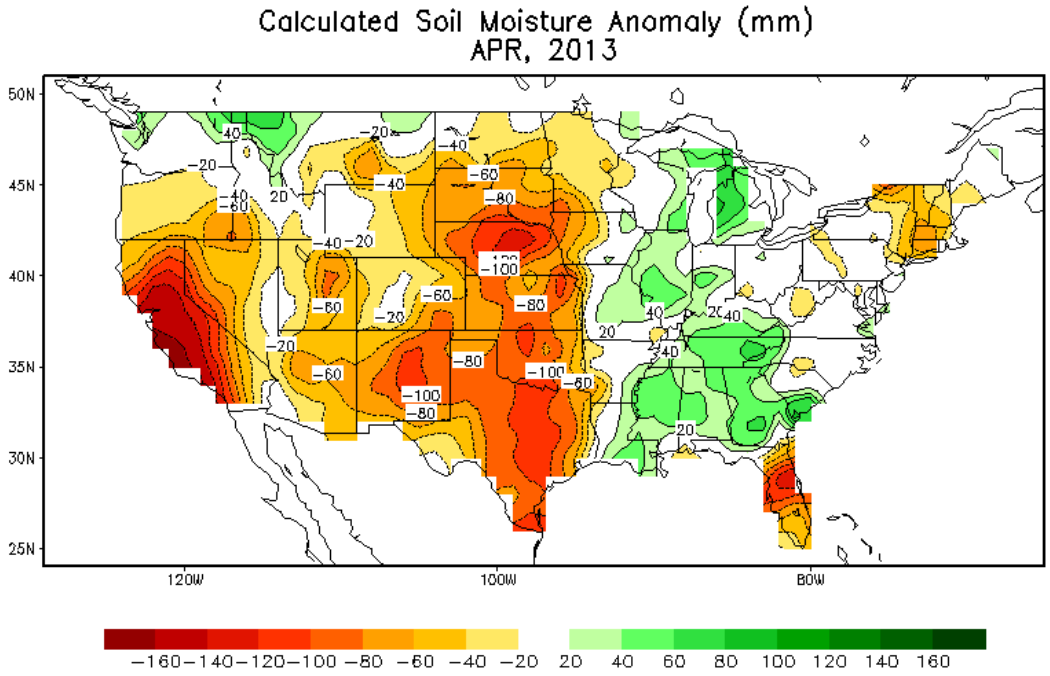


Single member one-month lag Precip-Temp correlation averaged for April – Aug base month (May – Sept temperature)

Soil Moisture as Initial Condition in NMME

- SM is one of the 5 extra phase I variables. SM is advertised to matter for seasonal prediction in the warm half of the year.
- Have yet to agree on units in practice
- Real time display in place (May starts)
- IC=about May 1, 2013.
- Forecast refers to June 2013

- Van den Dool previously discussed in NMME phone conference the strength of land-surface feedback as per P,T correlations in the model-world in the hindcast. Now real-time issues, AND soil moisture directly.

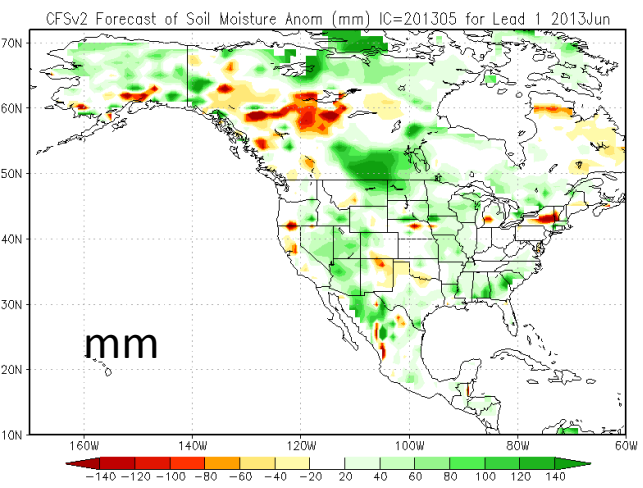


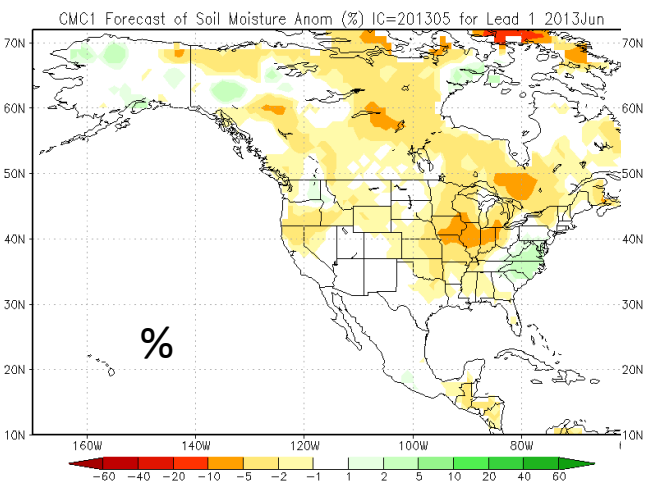
We borrow the color coding from CPC Leaky Bucket's SM.

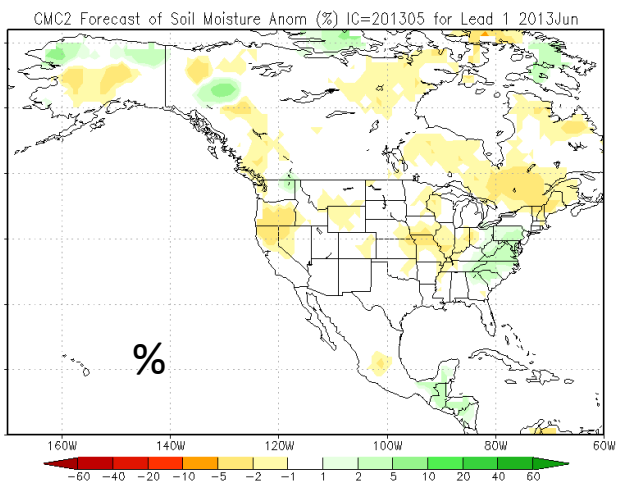
These are anomalies in mm. Anomaly is a matter of historical perspective. Climo=1981-2010. Typically ~50mm.

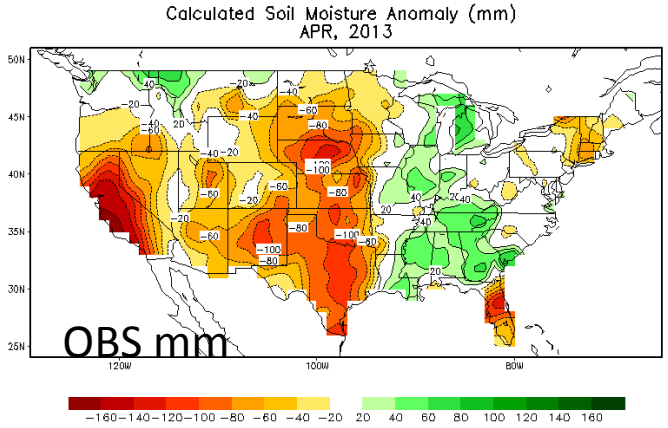
Six models in order:

- CFSv2
- CMC1
- CMC2
- GFDL
- NCAR
- NASA

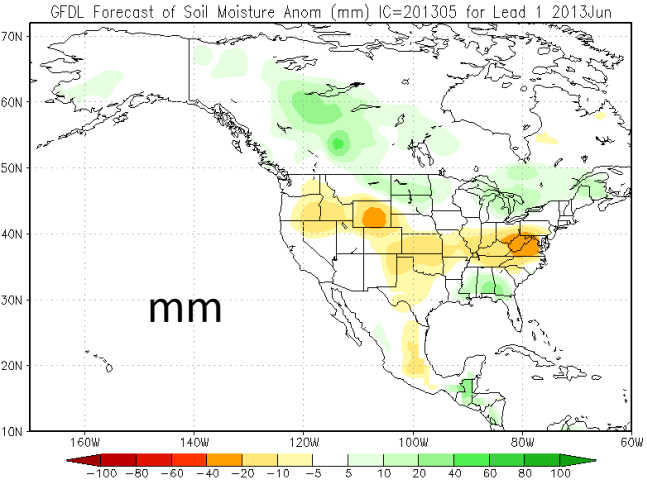


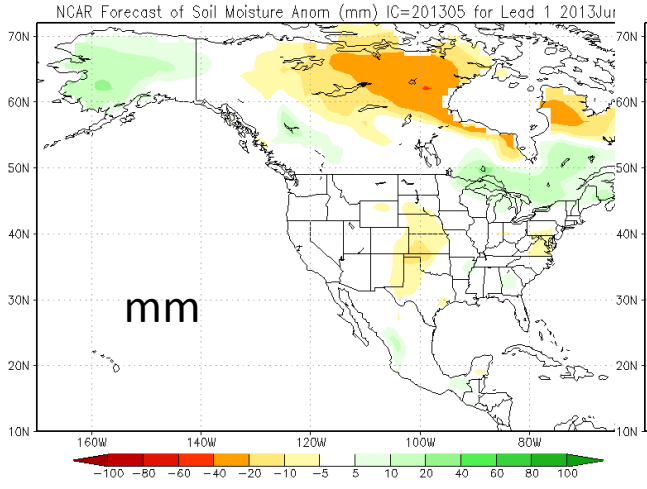


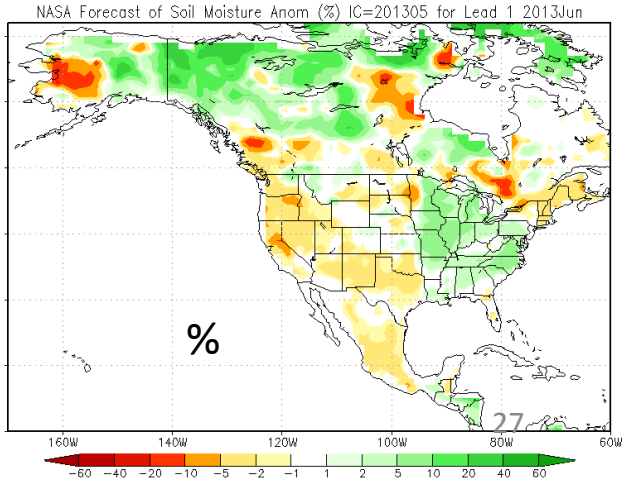




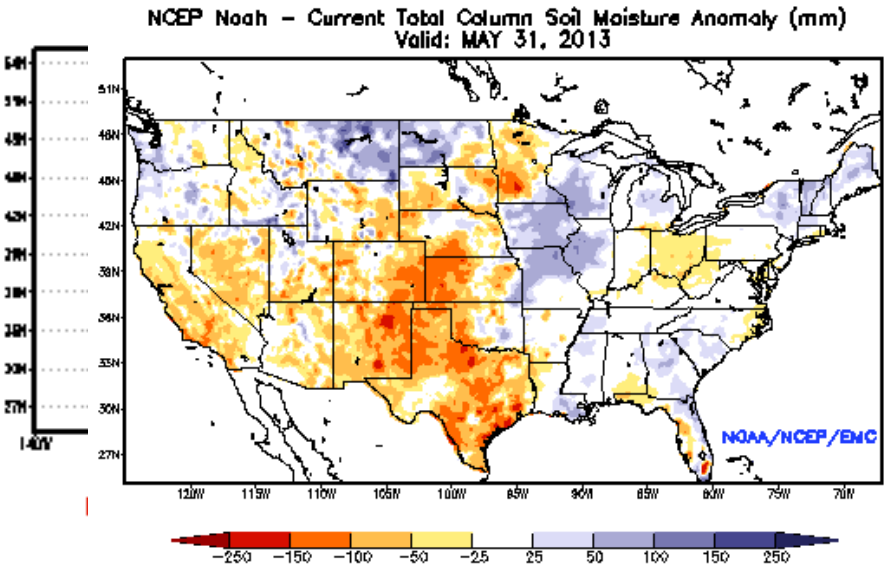
Units vary. Only NASA looks realistic.

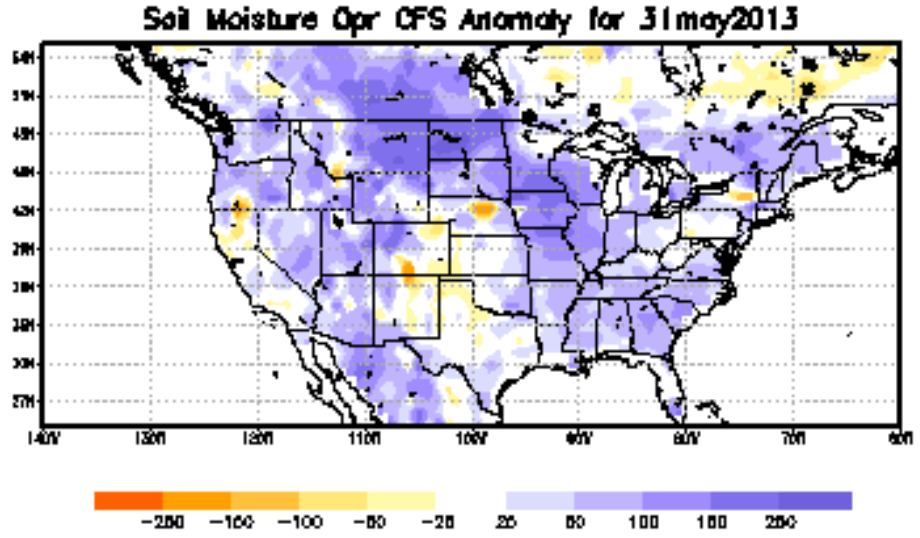


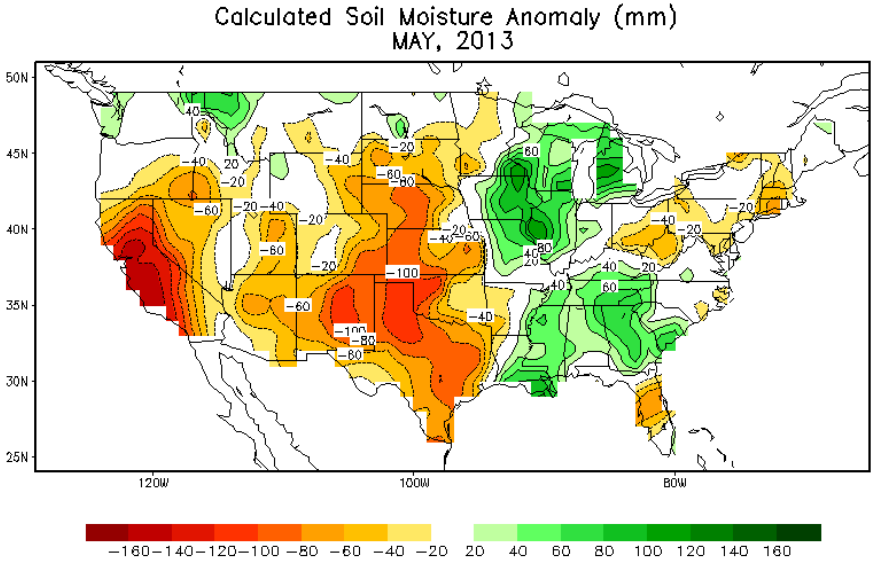




2013


NCEP Noah – Current Total Column Soil Moisture Anomaly (mm)
Valid: MAY 31, 2013
NOAA/NCEP/EMC
−250 −150 −100 −50 −25 25 50 100 150 250


Soil Moisture Opr CFS Anomaly for 31may2013
−250 −150 −100 −50 −25 25 50 100 150 250


Calculated Soil Moisture Anomaly (mm)
MAY, 2013
50N
45N
40N
35N
30N
25N
120W
100W
80W
−160 −140 −120 −100 −80 −60 −40 −20 20 40 60 80 100 120 140 160

Homogeneous and Heterogeneous Predictability and Forecast Skill in MME

Huug van den Dool, Emily Becker and Malaquias Pena

Acknowledgement: Other team members: Suranjana Saha, Peitao Peng
and all data suppliers (NCAR, GFDL, NASA, IRI etc), funding agents (CPO etc)

TMP2m Northern Hemisphere Leads 1-3

	cfsv1	cfsv2	echama	echamf	gfdl	nasa	ncar	obs (EM skill)	EM RMSE (C)	*EM SD*
cfsv1 EM	0.19	0.08	0.05	0.06	0.07	0.09	0.04	0.06	2.07	*0.814*
cfsv2 EM	0.09	0.27	0.09	0.08	0.16	0.19	0.01	0.19	1.98	*0.772*
echama EM	0.04	0.08	0.15	0.16	0.08	0.08	0.05	0.08	2.06	*0.765*
echamf EM	0.06	0.07	0.16	0.15	0.08	0.08	0.05	0.07	2.07	*0.760*
gfdl EM	0.06	0.14	0.07	0.06	0.25	0.15	0.01	0.15	2.08	*1.050*
nasa EM	0.07	0.14	0.07	0.05	0.15	0.27	0.00	0.14	2.06	*0.933*
ncar EM	0.03	0.01	0.04	0.04	-0.01	0.00	0.12	-0.01	2.24	*1.071*
singmem & obs SD	*2.280*	*2.086*	*2.128*	*2.086*	*2.367*	*1.991*	*2.263*	*2.143*		

Real Time Operations (contd)

- IRI had two entries in the first year, Echam_a and Echam_f. They completed the first year.
- Serious budget issues, and serious personnel cuts forced IRI to withdraw. Two models gone.
- We introduced two Canadian models in the two IRI slots. Still have 7 models.
- Mid-stream YR2: CFSv1 was terminated. 6 models left.
- For research purposes, the two IRI hindcast data sets will continue to be useful.

Some interesting NMME quirks

Is NMME more than the sum of its parts ? Hopefully it is. But the smallest common denomination does play a role. For instance:

- The longest lead of the NMME tends to be the smallest of the maximum leads of any particular model.
- The period of systematic error correction tends to be the shortest common hindcast record. Same for the climate anomalies.
- The NMME defaults to undefined at a particular gridpoint, if only one model is undefined. Therefore land-sea masks (and lakes) used by individual models and their interpolation techniques have an impact, especially when the resident resolution is low.

NMME Hindcast Data Repository

- All 12 start months for GFDL, CFSv1, CFSv2, IRIa, IRIf were provided right at the start in Aug 2011.
- NCAR and NASA hindcasts were made on the fly for each month, as we went along Aug 2011 – July 2012.
- In YR2:…… In YR3:…..
- Every new month gets added to the repository.
- This data is now available to the public for research
- Free download from IRI. Courtesy US Govt.

http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/

And some more ……

CPC would hardly be interested in MME unless there is a strong ongoing real time component to it.

CPC-NCEP is not an archiving place. We only distribute real time data (OK a few exceptions). Involving IRI or NCDC as archive and data distribution is the logical thing but not without complications.

"MME is known to be better (than the individual component model) on account of offsetting errors and increasing ensemble spread." Quote from NMME-ESPC-IOC report
three elements :
-) offsetting errors (what type of errors? systematic errors?, still true after calibration by model, given hindcasts?)
-) do we need to increase ensemble spread in 2013? (Probably not for monthly mean…..)
-) doing better than individual model is a low hurdle. Moreover we do not do better than an individual model in some elements.

Opinion: I am not convinced that NMME is better than using, say, CFSv2 alone with purposely and reasonably perturbed physics (multi-model in house), which is far easier to set-up, control and maintain and would be very useful for the 7 year road towards developing CFSv3.

Neglected subject: the transition of real time forecast data into the hindcast data base. NCEP (or NOAA really) failed.

Research likes to swarm (burst) ensemble members from the 1st (nominally) of the month, while Operations (NCEP) likes to generate members thruout the month. Real time vs Hindcast!!

IMME= CFSv2 + EUROSIP MODELS

	NCEP/CFSv2	ECMWF	UKMET	METF
Atmospheric Model	T126L64	Syst 4: T255L91	Glosea4 (120km) L85	T42L91 (T63-linear grid)
Ocean Model	MOM4 L40 0.25 deg Eq, 0.5 deg global	NEMO 0.3 deg Eq 1 deg global	NEMO L75 0.3 deg Eq 1 deg global	ORCA 0.5 deg Eq 2 deg global
Atmosphere/Ocean Coupling Frequency	30 minutes	3 hr	IN	IN
Land Model	NOAH 4-layer	IN	IN	IN
Sea Ice Model	3-layer interactive Seaice model	IN	IN	IN
Period of Hindcasts	1982-2010 (29 years)	1981-2010 (30 years)	1989-2002 (14 years)	1981-2009 (29 years)
Number of hindcast members	24(28)	15	12	11
Number of Leads	0-9 months	0-7 months	0-6 months	0-6 months

IN: Information needed

Intra-seasonal MME

You want to join MME????? OK, but

Basic Metrics and Protocol for CPC Forecast Tool Evaluations

June 2013 (Huang, Unger, van den Dool)

The metrics and protocol are intended to be
basic but broad enough to capture different aspects of evaluations,
for different types of forecasts and for different major climate henomena.

The focus is to evaluate the skills of new and experimental forecast tools
(both statistical and dynamic) for their impacts on CPC operations.

1. Basic Metrics

The metrics include different aspects of verifications including error/accuracy, skill scores, conditional statistics, reliability, biases, and etc. and three types of forecasts (namely, deterministic, categorical, and probabilistic). The metrics are consistent with those used in the CPC real-time verification system and at many places elsewhere.

Deterministic/Continuous:

- Anomaly Correlation (AC)
- Root Mean Square Error (RMSE)
- Mean Absolute Errors
- Amplitude
- Biases

Categorical:

- Contingency Table

Probabilistic:

- Brier Score (BS)
- Rank Probability Score (RPS)

Hindcast Protocol

The hindcast requirements mainly follow the current NMME protocol with some modification for sub-seasonal forecast. For testing potential use of new forecast tools for operational use, developers are required to provide hindcast data to CPC/CTB for systematic evaluations.

Hindcast start times must include all 12 calendar months

Hindcast period:

Monthly and seasonal forecasts: minimum 30 years (1982-2012)

Sub-seasonal forecasts: minimum 10 years (during 1999-2012)

Hindcast Protocol cntd

Forecast lead time:

Monthly and seasonal forecasts: 1- 7 months

Sub-seasonal forecasts: 1- 45 days

Number of ensemble members: free

Basic data:

Monthly and seasonal forecasts: Monthly mean of T2m, Prate and SST (or more)

Sub-seasonal forecasts: Daily data of OLR, U and V at 250 and 850 hPa, Z500, Z700, T2m Prate

Data distributed must include each ensemble member and total uncorrected fields.

Data format: Grid 1x 1

For your one stop shopping for NMME and IMME products, visit

http://www.cpc.ncep.noaa.gov/products/NMME/

1800 jpegs a month!

Conclusions I

T2m, NH land 23N-75N, 1982-2010. All start months, leads 1-3 combined, No CV

1. Prediction Skill (AC) varies from -0.01 to +0.19

2. Predicted sd(of EM) ~ .75-1C, which is much smaller than verifying SD (2-2.4C) for individual realization.

3. Point 2 is a matter of taking the ensemble mean (EM), some sqrt(N) damping of noise while retaining alleged signal.

4. All models have about the right inter-annual sd (near 2C)! Not over or under-done. This is high praise, and very different from under-dispersive reputation as per Demeter.
5. Homogeneous predictability (no CV required) ranges from 0.12 to 0.27. This is higher than skill reported in item 1 (0-0.19) but not hugely so.

6 For those who like high (homogeneous) predictability, there is not much to pick from among the 7 opinions.

Conclusions II

7. Heterogeneous predictability ranges from 0.00 to 0.19.
   Curiously this is already realized (item 1). Are better days ahead????

8. Symmetry by and large in "to predict" or "be predicted“

9. NCAR has a hard time predicting other models, or, be predicted by other models.
This is not bad in and of itself since we like orthogonal, but in this case.....

10. CFSv2, GFDL and NASA correlate the most to each other, AND,
    have the higher observed skill

11. The two IRI models predict each other,
   so maybe it is one ensemble of 24 as opposed to two ensembles of 12 each

12. In spite of shared pedigree CFSv1 and v2 do not predict each other very well.

About prate: -) all AC's are lower, but > 0 (less variation, greater dof than T2m),
-) homog: 0.06-0.15, heterog 0.01-0.06 (leaving IRI aside), obs skill 0.01-0.07